Full wwPDB X-ray Structure Validation Report (i)

Feb 1, 2016 – 06:42 PM GMT

PDB ID	:	4MH7
Title	:	Crystal structure of the catalytic domain of the proto-oncogene tyrosine-protein kinase MER in complex with inhibitor UNC1896
Authors	:	Zhang, W.; McIver, A.; Stashko, M.A.; Deryckere, D.; Branchford, B.R.; Hunter, D.; Kireev, D.B.; Miley, M.J.; Norris-Drouin, J.; Stewart, W.M.; Lee, M.; Sather, S.; Zhou, Y.; DiPaola, J.A.; Machius, M.; Janzen, W.P.; Earp, H.S.; Graham, D.K.; Frye, S.; Wang, X.
Deposited on	:	2013-08-29
Resolution	:	2.51 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.
We welcome your comments at validation@mail.wwpdb.org
A user guide is available at
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

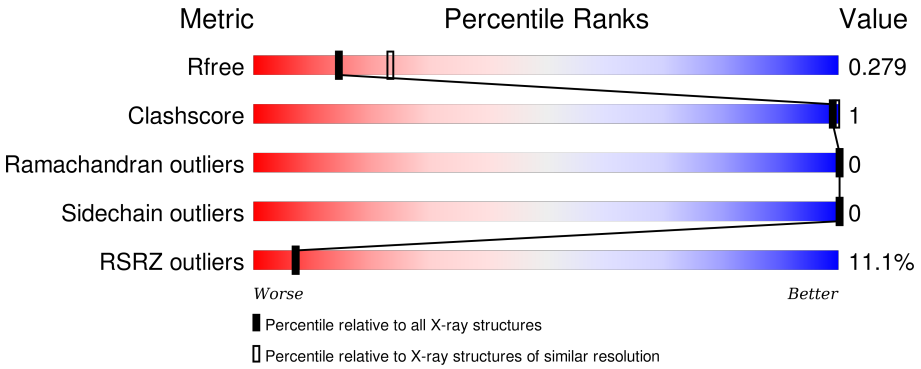
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.51 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	313	10%, 78%, 21%
1	B	313	8%, 81%, 19%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	CL	B	901	-	-	X	-

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 8198 atoms, of which 4088 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Tyrosine-protein kinase Mer.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	248	Total	C	H	N	O	S	0	0	0
			3983	1276	1995	331	362	19			
1	B	255	Total	C	H	N	O	S	0	0	0
			4114	1316	2058	342	379	19			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	552	MET	-	EXPRESSION TAG	UNP Q12866
A	553	GLY	-	EXPRESSION TAG	UNP Q12866
A	554	SER	-	EXPRESSION TAG	UNP Q12866
A	555	SER	-	EXPRESSION TAG	UNP Q12866
A	556	HIS	-	EXPRESSION TAG	UNP Q12866
A	557	HIS	-	EXPRESSION TAG	UNP Q12866
A	558	HIS	-	EXPRESSION TAG	UNP Q12866
A	559	HIS	-	EXPRESSION TAG	UNP Q12866
A	560	HIS	-	EXPRESSION TAG	UNP Q12866
A	561	HIS	-	EXPRESSION TAG	UNP Q12866
A	562	SER	-	EXPRESSION TAG	UNP Q12866
A	563	SER	-	EXPRESSION TAG	UNP Q12866
A	564	GLY	-	EXPRESSION TAG	UNP Q12866
A	565	LEU	-	EXPRESSION TAG	UNP Q12866
A	566	VAL	-	EXPRESSION TAG	UNP Q12866
A	567	PRO	-	EXPRESSION TAG	UNP Q12866
A	568	ARG	-	EXPRESSION TAG	UNP Q12866
A	569	GLY	-	EXPRESSION TAG	UNP Q12866
B	552	MET	-	EXPRESSION TAG	UNP Q12866
B	553	GLY	-	EXPRESSION TAG	UNP Q12866
B	554	SER	-	EXPRESSION TAG	UNP Q12866
B	555	SER	-	EXPRESSION TAG	UNP Q12866
B	556	HIS	-	EXPRESSION TAG	UNP Q12866
B	557	HIS	-	EXPRESSION TAG	UNP Q12866
B	558	HIS	-	EXPRESSION TAG	UNP Q12866

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	559	HIS	-	EXPRESSION TAG	UNP Q12866
B	560	HIS	-	EXPRESSION TAG	UNP Q12866
B	561	HIS	-	EXPRESSION TAG	UNP Q12866
B	562	SER	-	EXPRESSION TAG	UNP Q12866
B	563	SER	-	EXPRESSION TAG	UNP Q12866
B	564	GLY	-	EXPRESSION TAG	UNP Q12866
B	565	LEU	-	EXPRESSION TAG	UNP Q12866
B	566	VAL	-	EXPRESSION TAG	UNP Q12866
B	567	PRO	-	EXPRESSION TAG	UNP Q12866
B	568	ARG	-	EXPRESSION TAG	UNP Q12866
B	569	GLY	-	EXPRESSION TAG	UNP Q12866

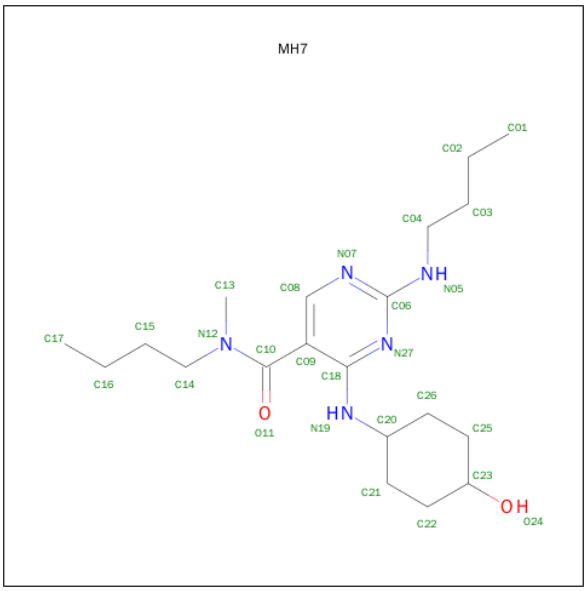
- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	4	Total Cl 4 4	0	0
2	A	4	Total Cl 4 4	0	0

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	1	Total Mg 1 1	0	0
3	A	1	Total Mg 1 1	0	0

- Molecule 4 is N-BUTYL-2-(BUTYLAMINO)-4-[(TRANS-4-HYDROXYCYCLOHEXYL)AMINO]-N-METHYLPYRIMIDINE-5-CARBOXAMIDE (three-letter code: MH7) (formula: $C_{20}H_{35}N_5O_2$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	A	1	Total	C	H	N	O	0	0
			62	20	35	5	2		

- Molecule 5 is water.

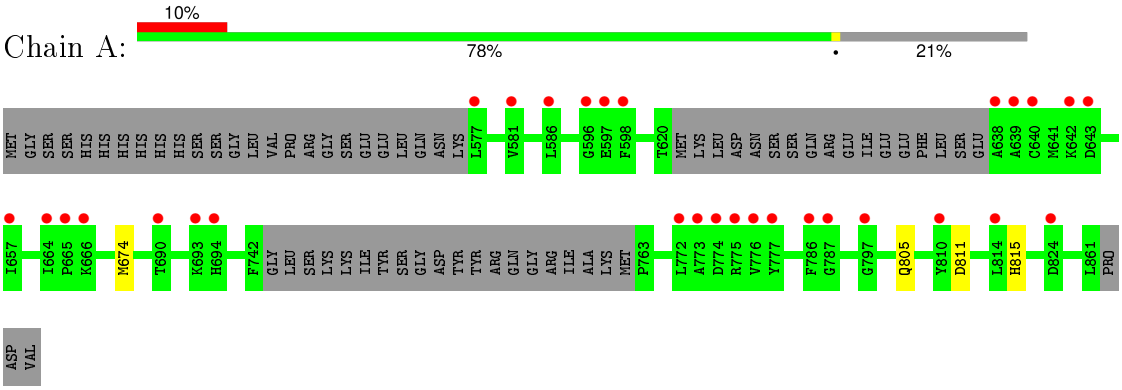
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	14	Total	O	0	0
			14	14		
5	B	15	Total	O	0	0
			15	15		

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

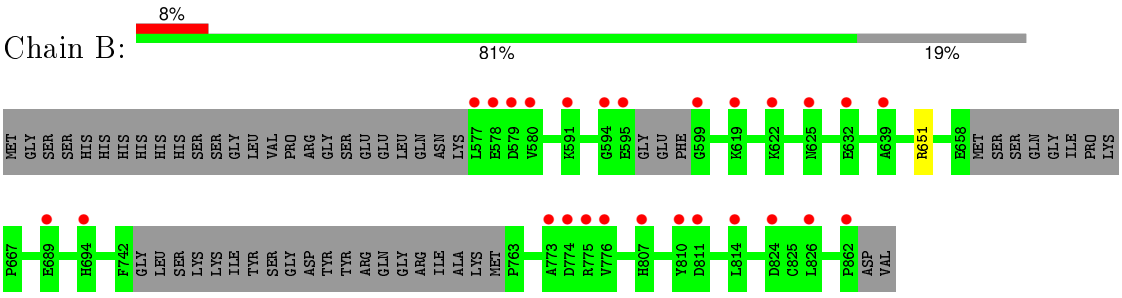
- Molecule 1: Tyrosine-protein kinase Mer

Chain A:



- Molecule 1: Tyrosine-protein kinase Mer

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	50.87Å 90.86Å 69.59Å 90.00° 99.98° 90.00°	Depositor
Resolution (Å)	32.06 – 2.51 32.06 – 2.51	Depositor EDS
% Data completeness (in resolution range)	83.0 (32.06-2.51) 80.2 (32.06-2.51)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.85 (at 2.51Å)	Xtriage
Refinement program	PHENIX (phenix.refine: dev_1468)	Depositor
R, R_{free}	0.223 , 0.279 0.238 , 0.279	Depositor DCC
R_{free} test set	1773 reflections (11.51%)	DCC
Wilson B-factor (Å^2)	29.6	Xtriage
Anisotropy	0.539	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.43 , 32.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Outliers	0 of 17725 reflections	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	8198	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.57% of the height of the origin peak. No significant pseudotranslation is detected.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MH7, MG, CL

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\| > 5$	RMSZ	$\#\|Z\| > 5$
1	A	0.21	0/2030	0.37	0/2744
1	B	0.21	0/2097	0.35	0/2832
All	All	0.21	0/4127	0.36	0/5576

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1988	1995	2002	3	0
1	B	2056	2058	2064	2	0
2	A	4	0	0	0	0
2	B	4	0	0	2	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
4	A	27	35	35	2	0
5	A	14	0	0	1	0
5	B	15	0	0	0	0
All	All	4110	4088	4101	6	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including

hydrogen atoms). The all-atom clashscore for this structure is 1.

All (6) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:651:ARG:NH1	2:B:901:CL:CL	2.69	0.62
1:A:674:MET:O	4:A:906:MH7:H131	2.02	0.58
1:A:811:ASP:O	1:A:815:HIS:ND1	2.39	0.55
4:A:906:MH7:O11	4:A:906:MH7:N19	2.38	0.52
1:A:805:GLN:NE2	5:A:1012:HOH:O	2.52	0.42
1:B:651:ARG:HG2	2:B:901:CL:CL	2.58	0.41

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	242/313 (77%)	235 (97%)	7 (3%)	0	100	100
1	B	247/313 (79%)	244 (99%)	3 (1%)	0	100	100
All	All	489/626 (78%)	479 (98%)	10 (2%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	222/280 (79%)	222 (100%)	0	100	100
1	B	231/280 (82%)	231 (100%)	0	100	100
All	All	453/560 (81%)	453 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 11 ligands modelled in this entry, 10 are monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
4	MH7	A	906	-	28,28,28	2.47	4 (14%)	30,36,36	2.15	8 (26%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
4	MH7	A	906	-	-	0/21/31/31	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	A	906	MH7	C09-C10	2.55	1.54	1.50
4	A	906	MH7	C06-N05	6.77	1.45	1.34
4	A	906	MH7	C10-N12	7.10	1.47	1.34
4	A	906	MH7	C18-N19	7.71	1.47	1.35

All (8) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	A	906	MH7	N07-C06-N27	-5.79	120.50	126.67
4	A	906	MH7	O11-C10-N12	-4.90	116.13	122.36
4	A	906	MH7	C18-C09-C10	-4.88	116.61	122.33
4	A	906	MH7	O11-C10-C09	-3.43	113.15	120.14
4	A	906	MH7	C04-N05-C06	-2.58	118.92	123.80
4	A	906	MH7	C09-C08-N07	-2.57	120.89	124.50
4	A	906	MH7	C06-N27-C18	2.23	121.49	116.92
4	A	906	MH7	C08-N07-C06	2.55	120.82	115.95

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	906	MH7	2	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	248/313 (79%)	0.90	30 (12%)	6	5	20, 31, 53, 70	0
1	B	255/313 (81%)	0.74	26 (10%)	9	9	20, 30, 50, 67	0
All	All	503/626 (80%)	0.82	56 (11%)	7	7	20, 31, 52, 70	0

All (56) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	579	ASP	7.2
1	A	598	PHE	6.3
1	A	775	ARG	6.3
1	B	775	ARG	5.4
1	A	664	ILE	5.2
1	A	640	CYS	5.0
1	B	810	TYR	4.8
1	A	824	ASP	4.7
1	B	595	GLU	4.5
1	A	773	ALA	4.3
1	A	639	ALA	4.0
1	B	773	ALA	4.0
1	A	597	GLU	3.9
1	B	862	PRO	3.8
1	A	643	ASP	3.7
1	A	776	VAL	3.6
1	B	814	LEU	3.6
1	A	596	GLY	3.5
1	A	657	ILE	3.4
1	A	774	ASP	3.4
1	A	581	VAL	3.3
1	B	774	ASP	3.2
1	B	632	GLU	3.2
1	A	693	LYS	3.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	807	HIS	3.1
1	B	811	ASP	3.0
1	B	578	GLU	3.0
1	B	776	VAL	3.0
1	A	777	TYR	2.9
1	A	772	LEU	2.9
1	A	665	PRO	2.8
1	B	599	GLY	2.8
1	A	666	LYS	2.8
1	B	639	ALA	2.6
1	B	619	LYS	2.5
1	A	814	LEU	2.5
1	B	577	LEU	2.5
1	A	577	LEU	2.5
1	A	642	LYS	2.4
1	B	625	ASN	2.4
1	A	586	LEU	2.4
1	B	591	LYS	2.4
1	A	786	PHE	2.3
1	B	622	LYS	2.3
1	A	694	HIS	2.3
1	B	689	GLU	2.3
1	B	594	GLY	2.3
1	A	810	TYR	2.3
1	B	826	LEU	2.2
1	B	694	HIS	2.2
1	A	797	GLY	2.2
1	A	690	THR	2.1
1	B	580	VAL	2.0
1	A	638	ALA	2.0
1	B	824	ASP	2.0
1	A	787	GLY	2.0

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
4	MH7	A	906	27/27	0.91	0.24	0.73	28,37,45,47	0
3	MG	B	905	1/1	0.89	0.25	0.22	34,34,34,34	0
2	CL	A	902	1/1	0.99	0.22	0.14	24,24,24,24	0
2	CL	B	902	1/1	0.99	0.14	-0.87	24,24,24,24	0
2	CL	B	901	1/1	0.91	0.11	-1.62	34,34,34,34	0
2	CL	A	901	1/1	0.99	0.12	-2.28	24,24,24,24	0
2	CL	B	904	1/1	0.93	0.08	-	47,47,47,47	0
2	CL	A	903	1/1	0.97	0.11	-	52,52,52,52	0
2	CL	A	904	1/1	0.95	0.08	-	29,29,29,29	0
3	MG	A	905	1/1	0.73	0.14	-	41,41,41,41	0
2	CL	B	903	1/1	0.99	0.14	-	43,43,43,43	0

6.5 Other polymers (i)

There are no such residues in this entry.